AutoML Applications for Bacilli Recognition by Taxonomic Characteristics Determination over Microscopic Images

Aleksei Samarin, Aleksei Toropov,
Alina Dzestelova, Artem Nazarenko,
Egor Kotenko, Elena Mikhailova,
Valentin Malykh
ITMO University
St. Petersburg, Russia
avsamarin@itmo.ru, toropov.ag@hotmail.com,
aldzestelova@gmail.com, aanazarenko@itmo.ru,
kotenkoed@gmail.com, e.mikhailova@itmo.ru

Alexander Savelev, Alexander Motyko
St. Petersburg Electrotechnical University “LETI”
St. Petersburg, Russia
algsavelev@gmail.com, aamotyko@etu.ru

Aleksandra Dozortseva
St. Petersburg State Institute of Technology
(Technical University)
St. Petersburg, Russia
adozorceva@rambler.ru

***Abstract*—In this work, we describe our research aimed at developing classifiers for microbial images (bacilli images) obtained through microscopy of live (non-static) samples. We employed our proposed approach called AutoML, which is based on the automatic generation and analysis of the feature space to create the most optimal descriptors for microscopic images used in their classification. This approach allows us to utilize interpretable taxonomic features based on the external geometric characteristics of images of various types of microorganisms. To demonstrate the effectiveness of our proposed solution, we also publish an annotated dataset we collected, containing microbial images of unfixed microscopic scenes. Additionally, we compare the classification performance of our solution with the results of various types of classifiers, including those based on deep neural network models. Our approach showed the best results among those studied (Precision = 0.989, Recall = 0.992, F1-score = 0.990).**

I. INTRODUCTION

In the domain of computer vision, numerous challenges associated with categorizing images possessing distinct visual attributes are highly pertinent [1]–[13]. Although contemporary neural network classifiers exhibit exceptional proficiency in addressing the classification of images featuring diverse objects against a natural backdrop [14]–[26], there exist several categories of images whose visual semantics diverge significantly from the geometric primitives observed in general object images [1]–[13]. Notable among these categories are images depicting scenes with graphic text elements [1]–[7], [9], various mechanical images, and importantly, biomedical images [8], [10]–[13], specifically those acquired through microscopy [27]–[29]. This study is also focused on developing methodologies for classifying images derived from microscopy.

It is important to recognize that the automation of microorganism image classification has extensive practical implications [27]–[32]. For instance, it facilitates the automation of laboratory biomedical analyses of various patient samples, special control examinations of swabs from different food products and raw materials for sanitary inspection, and the evaluation of numerous water samples from tanks, swimming pools, natural bodies of water, etc. Typically, during microscopic examinations, the studied scene is pre-fixed and stained to enhance the visual characteristics of the scene being analyzed. However, to save time, particularly on an industrial scale, it is crucial to conduct microscopic studies on unfixed and unstained scenes. In this context, visual analysis, and consequently its automation, becomes considerably more complex. The complexity arises from artifacts that occur under conditions of an unfixed scene and the potential movement of microorganisms, such as blurred boundaries, unclear edges, and insufficient visibility of certain parts of the object being examined. Similar issues that hinder visual analysis also emerge when the scene is unstained. Consequently, additional features, along with the specific semantics of geometric primitives, make the classification of images more challenging.

A wide range of approach families were examined, consistent with numerous studies that focus on images with particular visual characteristics to determine the best methods for building classifiers [1]–[13].

Deep neural network (DNN) models were inevitably included among the various method groups studied [14]–[17], [20]–[22], [25], [26]. For a considerable period, this family of approaches has been the frontrunner in the classification of both general images and numerous other image types. Since the rapid rise in popularity of deep neural network architectures, their designs and operational principles have seen substantial changes. Today, there are several common categories of neural network methods, each distinguished by specific features.

We first analyzed well-established convolutional models as our initial group of neural network classifiers. These models were instrumental in bringing neural networks to the forefront in classifying diverse objects. Numerous approaches, such as [20]–[22], [33], are currently available. These methods achieve high performance in classifying general objects across many classes. However, purely convolutional networks have limitations, such as localized feature extraction at each level of abstraction and the lack of explicit mechanisms for interpreting feature space elements, which significantly restrict their application in biomedical data analysis. Additionally, this class of models typically requires substantial amounts of training data, especially for specialized domains.

Various implementations of the attention mechanism were introduced during the enhancement of deep neural network classifiers, enabling the circumvention of limitations related to feature locality in convolutional neural networks. Currently, self-attention is the most prevalent type of attention in deep neural network encoders, with non-local blocks being their basic form [34], [35]. The development of non-local blocks led to the creation of multi-headed self-attention, which is foundational to the transformer architecture [18]. Transformer-based architectures now dominate many computer vision tasks, achieving top rankings in classification tasks [23], [24]. However, certain characteristics of this neural network family are noteworthy. These classifiers are generally trained on extensive datasets and can be sensitive to the scale of objects within the scene, limiting their direct application to specific domains like microscopic images. Moreover, using this type of neural network as an encoder in classification models does not address the challenges of interpretability and the creation of an analytical description of domain-specific entities and features.

In addition to advancements in structural components, the development of intermediate representations and hidden spaces in neural networks has also made significant progress. The introduction of CLIP and BLIP family architectures [15]–[17] combined the semantics of visual scenes with their textual descriptions within a unified embedding space. This innovation has greatly improved scene interpretability and the zero-shot paradigm, mainly aiding domains with heterogeneous objects in natural backgrounds, rather than specific domains. It is also crucial to note that these architectures necessitate large training datasets and wide coverage, which is difficult to achieve in specialized fields.

In the context of image classification, the analytical specification of procedures for calculating image descriptors is essential for the unambiguous interpretation of feature spaces [36]–[39]. A prevalent method includes histogram descriptors derived from analyzing histograms of oriented gradients [36], [37] and local binary patterns [38], [39]. While these vectorization techniques are not as powerful as the neural network approaches for classifying extensive sets of diverse objects, they provide inherent interpretability related to fundamental geometric characteristics.

Specific geometric characteristics, beyond the usual descriptors based on histogram features, are also significant. These characteristics highlight the physical properties of the objects' shapes, which is particularly important in taxonomy. Additionally, to develop the most optimal configurations of analytically defined features, including their generalizations and associated parameters, it is beneficial to use AutoML and automatic feature generation techniques.

In this study, we tackle the challenge of classifying bacilli microorganisms from microscopy images. We devised a method leveraging AutoML techniques, built upon analytically defined methods for calculating visual features, to achieve interpretable object characteristics and enhance the solution's taxonomic relevance. We assessed our proposed solution's efficiency by comparing it with the aforementioned image classification methods. To train, validate, and evaluate object properties, we compiled and annotated a new dataset, which is now publicly available.

II. Problem statemet

This research primarily focuses on the challenge of binary image classification. An input image is labeled as 1 if it contains a single bacilli instance; otherwise, it is assigned a label of 0 (see Fig. 1).

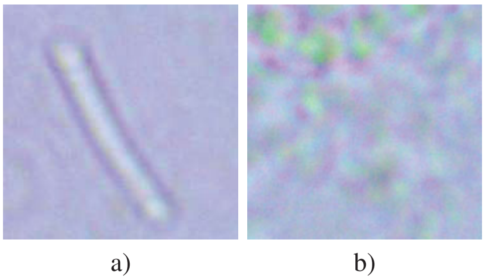

Fig. 1. Examples of images from the two classes: a) an image of bacilli; b) an image of a random region of microscopy scene that does not contain bacilli.

III. Proposed solution

The proposed method consists of a multi-phase pipeline. First, images are preprocessed to adjust visual attributes, enhancing classification accuracy. In the subsequent phase, features are extracted analytically, and aggregate features are generated automatically. The final classifier is then applied to the vectorized representation of the input image. Figure 2 illustrates the overall architecture of the classifier.

A. Image preprocessing

We utilized the model described in [40] for preprocessing, making several adjustments as specified below. The structure of the resulting corrective transformation diagram is as follows:

$$I_e = I_o + \sum_{i=1}^{n} f_i(I_o, h_i(I_{so})).$$

The structure consists of multiple independent blocks, with the number of blocks depending on the filters used. Each i-th block processes a scaled-down version of the original image I_{so} through the parameter generator h_i, which generates parameters p_i for the corresponding filter f_i. The filters are applied individually to the original image I_o, and the final

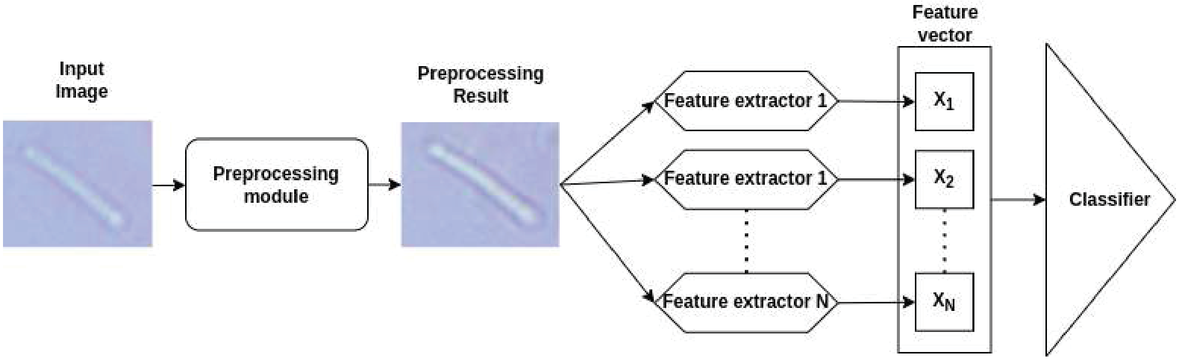

Fig. 2. The general structure of the proposed model for bacilli image classification

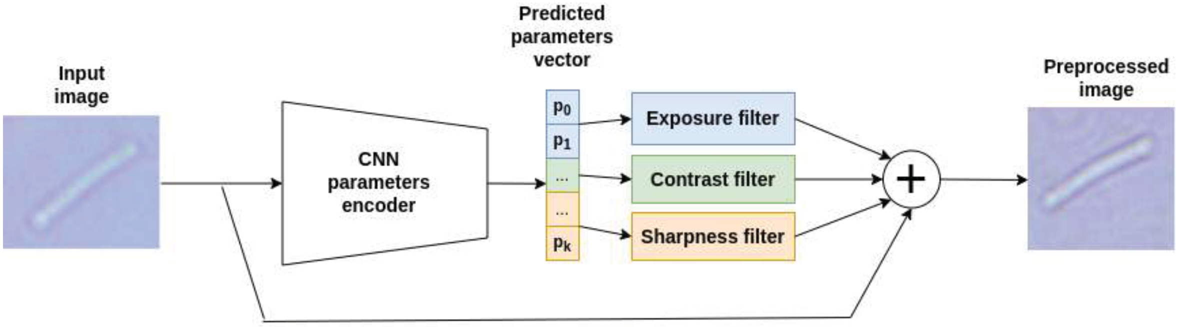

Fig. 3. Image preprocessing module structure

enhanced image is created by summing the original image and the outputs of the filters.

We employed filters designed for corrective transformations to address the earlier discussed characteristics of unfixed microscopic scene images, such as blurring, low contrast, and lack of sharpness.

The ***sharp*** filter is defined using auxiliary formula:

$$I_{out} = I_{in} \circledast \frac{1}{\nu}(K + M \cdot q),$$

where K – filter kernel matrix, M – map matrix with the same shape as K and ν is sum of elements of $(K + M \cdot q)$ for kernel matrix normalization. The formula mentioned above is independently applied to the red, green, and blue channels, each with its own trainable parameter. Consequently, the parameters for defining the sharp filter modification are as follows:

$$K = \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & -476 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix},$$

$$M = \begin{pmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.9 & 0.9 & 0.9 & 0.8 \\ 0.8 & 0.9 & 1 & 0.9 & 0.8 \\ 0.8 & 0.9 & 0.9 & 0.9 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{pmatrix},$$

Automatic ***contrast*** adjustment is achieved by manipulating $p \in [-1, 1]$, which specifies the transformation applied to each pixel of the input image. Thus, the original image undergoes the following mapping:

$$I_{out}[x, y] = \begin{cases} (I_{in}[x, y] - 0.5) \cdot \frac{1}{1-r}, & \text{if } r > 0 \\ (I_{in}[x, y] - 0.5) \cdot (1 - r), & \text{otherwise;} \end{cases}$$

It is also important to mention that general exposure adjustments are required due to the highly variable lighting conditions during microscopic examination.

The following image transformation performs automatic ***exposure*** correction:

$$I_{out}[x, y] = I_{in}[x, y] \cdot 2^t.$$

Because of the reduced number of transformations, we were able to integrate predictors for all parameters into a single neural network encoder.

Thus, the architecture of the preprocessing module is shown in Figure 3.

B. Automated feature generation

We analyze a wide range of diverse characteristics of microorganisms extracted from computer microscopy images in our approach. The initial step involves recording various parameters of the target object. For clarity, we have categorized the extracted characteristics and their handling principles into three main groups.

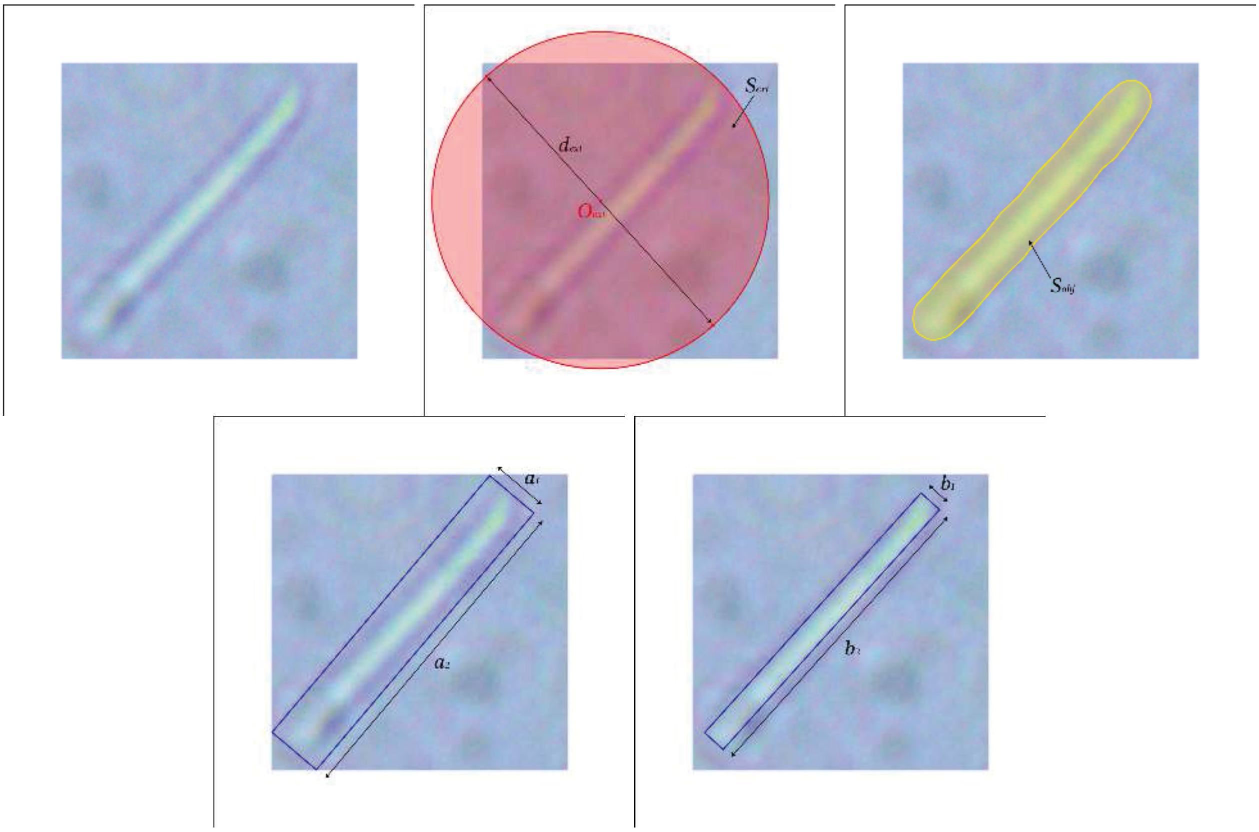

Fig. 4. The first group of the object's characteristics

1) The first group: In this context, we derive several evident characteristics from the obtained image related to the object's size or components. These characteristics include the diameter of the circumscribed circle d_{ext}, the area of the circumscribed circle S_{ext}, and the area of the object itself S_{obj}. Additionally, we consider the length and width of the minimum area rotated circumscribed rectangle — a_1 and a_2, respectively, and the length and width of the maximum area rotated inscribed rectangle — b_1 and b_2, respectively, along with numerous other attributes describing the investigated object. Some examples of the first group features are shown in Figure 4.

We further document different pairs of interrelated parameters and examine the potential values of their ratios. By doing so, we create a set of numbers β_0, where each element is the ratio of one characteristic to another, with the two being connected in some way. For instance, we can assume that

$$\beta_{00} = \frac{d_{ext}}{a_1},$$

$$\beta_{01} = \frac{S_{ext}}{a_1 \cdot a_2},\ \beta_{02} = \frac{S_{obj}}{a_1 \cdot a_2},\ \beta_{03} = \frac{S_{obj}}{b_1 \cdot b_2},$$

$$\beta_{04} = \frac{a_1}{a_2},\ \beta_{05} = \frac{b_1}{b_2},\ \beta_{06} = \frac{b_2}{a_2},$$

and so on.

To identify the defining characteristics of microscopic objects and further classify various microorganisms, we can explicitly utilize all values from this β_0 set. These characteristics can be derived not only from a sample of explicit values of previously extracted parameter ratios β_{0i} or the products of these ratios with some experimentally selected numerical coefficients $\alpha_{0i}\beta_{0i}$, but also from their various linear combinations in the form

$$\sum \gamma_{0j} \sum {\alpha_0}_i^j {\beta_0}_i^j * 1_{A_k}(j),$$

where γ_{0j} represents additional significant numerical coefficients obtained as a result of the training process, $A_k \in 2^{\{1,\ldots,i*j\}}$, and $k \in [1..i*j]$.

2) The second group: In this scenario, compared to the previous section, we extract less obvious characteristics of the object under examination. These include, for instance, the maximum L_{max} and minimum L_{min} distances from the object's center of mass to its contour, and the radius of the circumscribed circle r_{ext}. Additionally, we consider the distance $dist(O_m; O_{ext})$ between the center of mass O_m and the center of its circumscribed circle O_{ext}, the distance $dist(O_m; O_{rec}ext)$ between the center of mass O_m and the center of its maximum area rotated circumscribed rectangle $Orecext$, and the distance $dist(O_m; Orecint)$ between the center of mass O_m and the center of its maximum area rotated inscribed rectangle $Orec_{int}$, among others. Examples of features extractable within this second group are illustrated in Figure 5.

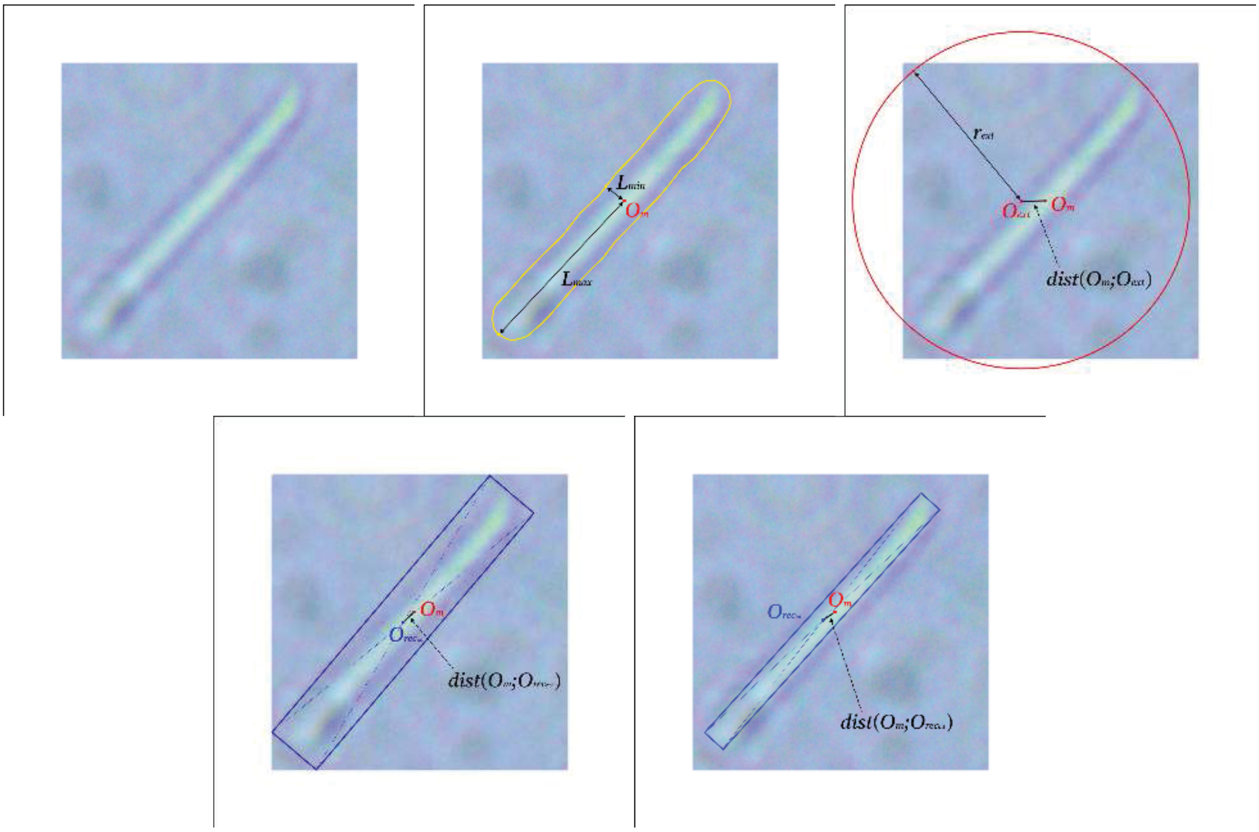

Fig. 5. The second group of the object's characteristics

Additionally, we document diverse pairs of correlated parameters and explore the potential values of their proportions. Consequently, we establish a numerical set β_1, where each member represents a ratio of one previously discussed characteristic's magnitude to another, somehow linked to the initial characteristic. For instance, we assume that

$$\beta_{10} = \frac{L_{max}}{L_{min}},$$

$$\beta_{11} = \frac{dist(O_m; O_{ext})}{r_{ext}}, \; \beta_{12} = \frac{dist(O_m; O_{ext})}{L_{max}},$$

$$\beta_{13} = \frac{dist(O_m; O_{recext})}{r_{ext}}, \; \beta_{14} = \frac{dist(O_m; O_{recint})}{L_{min}},$$

and so forth.

Next, we can use all the values from this β_1 set to explicitly identify the distinctive features of microscopic entities and accurately classify various microorganisms. The characteristics of interest can be constructed not only from a sample of explicit values of previously obtained parameter ratios β_{1i} or from the products of these ratios with some empirically chosen numerical coefficients $\alpha_{1i}\beta_{1i}$, but also their different linear combinations of the form

$$\sum \gamma_{1j} \sum \alpha_{1}{}_{i}^{j} \beta_{0}{}_{i}^{j} * 1_{A_k}(j),$$

where γ_{1j} represents additional significant numerical coefficients obtained as a result of the training process, $A_k \in 2^{\{1,\ldots,i*j\}}$, and $k \in [1..i*j]$.

3) The third group: This group encompasses numerous parameters derived from images of microorganisms through microscopy, mainly associated with various geometric objects. We consider the following: the distance K_1max between the center of its circumscribed circle O_{ext} and the center of its maximum area rotated circumscribed rectangle $Orecext$, the distance K_2max between the center of its circumscribed circle O_{ext} and the center of its inscribed rectangle $Orecint$, the maximum distance from the object's contour to its maximum area rotated circumscribed rectangle $K_1recmax$ and to its inscribed rectangle $K_2recmax$, as well as several other characteristics defining the examined object. Examples of features extracted in this third group are shown in Figure 6.

Additionally, we record a range of interdependent parameter pairs and examine the possible values of their ratios. This process results in the creation of a numerical set β_2, where each element signifies the ratio of one object characteristic to another, connected to the primary attribute. For instance, we assume that

$$\beta_{20} = \frac{K_{1max}}{K_{2max}}, \; \beta_{21} = \frac{K_{2max}}{K_{1max}},$$

$$\beta_{22} = \frac{K_1rec_{max}}{K_2rec_{max}}, \; \beta_{23} = \frac{K_2rec_{max}}{K_1rec_{max}},$$

and more.

Subsequently, all the values from the β_2 set can be utilized to determine the unique characteristics of microscopic organ-

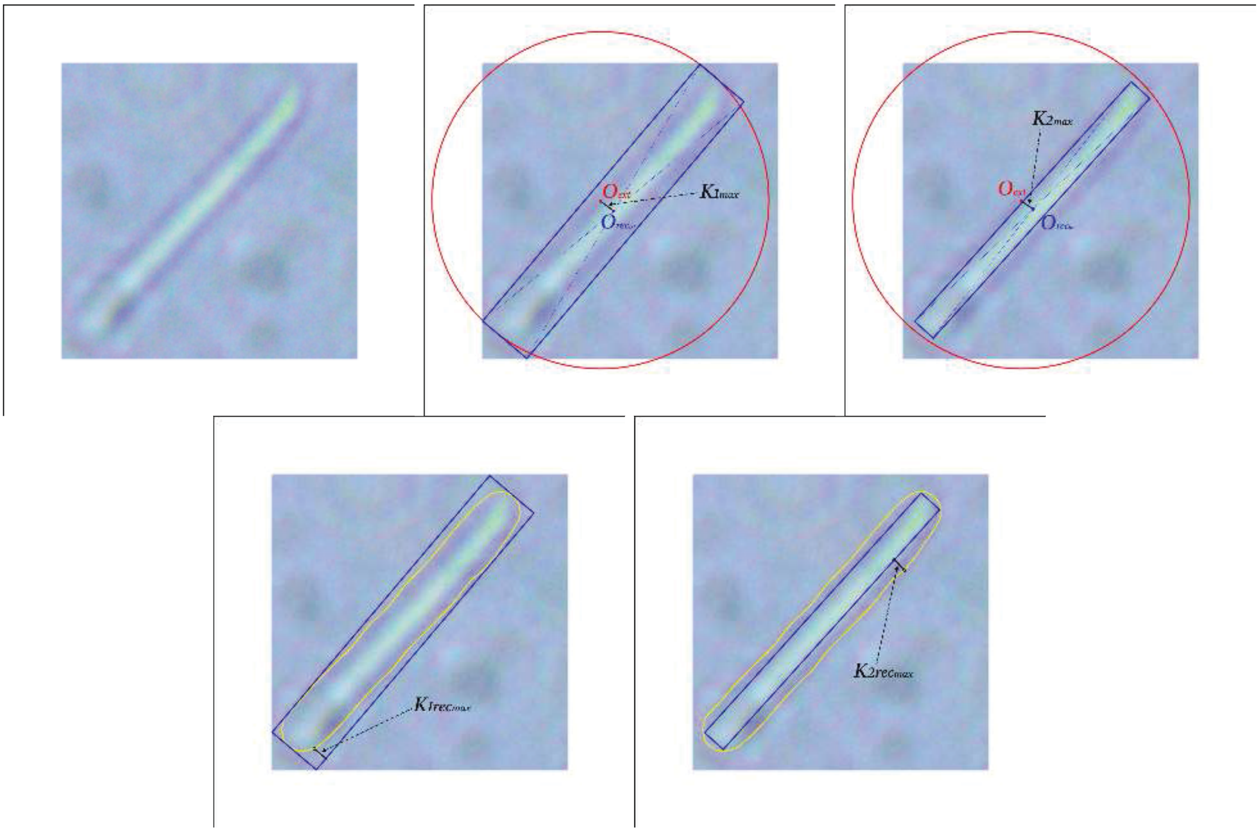

Fig. 6. The third group of the object's characteristics

isms and to classify various obtained objects. The defining parameters of each object can be constructed not only from the set of ratios β_{2i} mentioned earlier or from the products of these ratios with certain experimentally chosen coefficients α_{2i}, but also from various linear combinations, as follows:

$$\sum \gamma_{2j} \sum \alpha_{2}{}_i^j \beta_{2}{}_i^j * 1_{A_k}(j),$$

where γ_{2j} represents additional significant numerical coefficients obtained as a result of the training process, $A_k \in 2^{\{1,\ldots,i*j\}}$, and $k \in [1..i*j]$.

C. Classifiers

For classification purposes, we employed various types of classifiers commonly utilized in vector space classification tasks. The methods explored included Support Vector Machines (SVM) [41], Linear Regression (LR) [42], Random Forests (RF) [43], Gradient Boosting Machines (GBM) [44], and Fully Connected Neural Networks (FCN) [45]. As evidenced in the experiments section, the GBM classifier achieved the highest performance within our combined classifier.

IV. Evaluation

A. Dataset description

We collected and published our own dataset to train our classifier, assess its effectiveness, and compare it with other models. The dataset comprises images of target microorganisms, other types of microorganisms, areas devoid of microorganisms, and random fragments of microscopic scenes obtained during product microscope examinations (Figure 7).

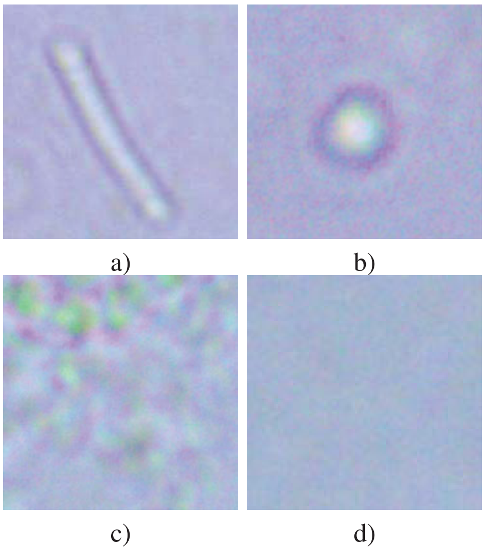
a) b) c) d)

Fig. 7. *Examples of images from the proposed dataset (MBID): a) an image of bacilli; b) an image of other microbial; c) an image of a random region of microscopy scene that does not contain bacilli; d) an image of a region of microscopy scene without any microorganisms.*

The complete dataset comprises approximately 5000 im-

TABLE I. COMPARATIVE ANALYSIS OF CLASSIFICATION PIPELINES USING THE MMICD DATASET (TOP-30)

Filters configuration	Classifier model	Precision	Recall	F1
Exposure + Contrast + Sharpness	MobileNetV3	0.887	0.889	0.888
Exposure + Contrast + Sharpness	InceptionResNetV1	0.891	0.893	0.892
Exposure + Contrast	ResNet152	0.893	0.896	0.894
Exposure + Contrast	EfficientNetB0	0.896	0.896	0.896
Exposure + Contrast	Generated features + SVM	0.901	0.899	0.900
Exposure + Contrast	InceptionResNetV2	0.901	0.901	0.901
Exposure + Contrast + Sharpness	Generated features + SVM	0.906	0.907	0.907
Exposure + Contrast + Sharpness	EfficientNetB0	0.909	0.910	0.909
Exposure + Contrast	ResNet101	0.912	0.915	0.913
Exposure + Contrast + Sharpness	EfficientNetB1	0.919	0.921	0.920
Exposure + Contrast	EfficientNetB2	0.924	0.928	0.926
Exposure + Contrast + Sharpness	ResNet101	0.929	0.935	0.932
Exposure + Contrast + Sharpness	EfficientNetB3	0.934	0.939	0.936
Exposure + Contrast	EfficientNetB4	0.939	0.942	0.940
Exposure + Contrast	CoAtNet	0.940	0.944	0.942
Exposure + Contrast	EfficientNetB6	0.941	0.949	0.945
Exposure + Contrast	Generated features + RF	0.944	0.953	0.948
Exposure + Contrast	SE-ResNext50	0.949	0.955	0.952
Exposure + Contrast + Sharpness	ResNet152	0.955	0.957	0.956
Exposure + Contrast + Sharpness	Generated features + RF	0.959	0.960	0.959
Exposure + Contrast	Generated features + GBM	0.961	0.960	0.960
Exposure + Contrast + Sharpness	CoAtNet	0.964	0.963	0.963
Exposure + Contrast	ViT-L/16	0.967	0.966	0.966
Exposure + Contrast	EfficientNetB3	0.968	0.969	0.968
Exposure + Contrast + Sharpness	EfficientNetB4	0.973	0.970	0.971
Exposure + Contrast + Sharpness	InceptionResNetV2	0.976	0.973	0.974
Exposure + Contrast + Sharpness	SE-ResNext50	0.978	0.977	0.977
Exposure + Contrast + Sharpness	ViT-L/16	0.983	0.981	0.982
Exposure + Contrast + Sharpness	EfficientNetB6	0.986	0.984	0.985
Exposure + Contrast + Sharpness	**Generated features + GBM**	**0.989**	**0.992**	**0.990**

ages, categorized into two classes and three groups for training, testing, and hyperparameter tuning. These groups contain around 2500, 250, and 250 images, respectively. The original images were captured using a Levenhuk MED D30T microscope. This dataset has been made publicly accessible as the Microscopy Bacilli Images Dataset [46].

B. Experiments

Thus, evaluating the values of ratios and their linear combinations allows us to identify various significant features of microscopic objects. This is essential for developing advanced machine-learning models that can accurately detect and classify microorganisms in images. We assessed the performance of the considered models and their configurations by evaluating the Precision, Recall, and F1-score of a set of classifiers.

Among the preview classifiers, we have selected groups of methods that have shown themselves to work best when processing biomedical images and microscope images. The considered neural network architectures have both convolutional building blocks and self-attention mechanisms for the most efficient extraction of characteristic features.

Additionally, we conducted an ablation study on different filter combinations of LFIEM [40] trained on our dataset to reconstruct the original image from its distorted version to enhance preprocessing for our combined classifier scheme. The top results are presented in Table I for reference.

As shown in Table I, the best results were achieved using our model with a preprocessing configuration that included exposure, contrast, and sharpness filters, along with generated feature extractors and GBM as the classifier.

V. CONCLUSION

We developed a hybrid neural network architecture for classifying unfixed bacilli microscopic images in our research. To assess the effectiveness of this method, we compiled, annotated, and publicly released a dedicated dataset. By leveraging explicitly specified interpretable taxonomic features, our approach constructs characteristic image descriptors, outperforming other methods on the tested dataset. Additionally, using our pipeline, we identified a set of interpretable taxonomic features (detailed in this paper) that can be employed independently of our classifier to manually identify microorganism types from microscopic images. In the future, we aim to apply these techniques to recognize other microorganisms and perform comprehensive analyses of microscopic scenes.

REFERENCES

[1] T. Tsai, W. Cheng, C. You, M. Hu, A. W. Tsui, and H. Chi, “Learning and recognition of on-premise signs from weakly labeled street view images,” *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1047–1059, March 2014.

[2] S. Romberg, L. G. Pueyo, R. Lienhart, and R. van Zwol, “Scalable logo recognition in real-world images,” in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ser. ICMR ’11. New York, NY, USA: ACM, 2011, pp. 25:1–25:8. [Online]. Available: http://www.multimedia-computing.de/flickrlogos/

[3] A. Samarin and V. Malykh, “Worm-like image descriptor for signboard classification,” in *Proceedings of The Fifth Conference on Software Engineering and Information Management (SEIM-2020)*, 2020.

[4] T. Chattopadhyay and A. Sinha, "Recognition of trademarks from sports videos for channel hyperlinking in consumer end," in *2009 IEEE 13th International Symposium on Consumer Electronics*, May 2009, pp. 943–947.

[5] A. Samarin, V. Malykh, and S. Muravyov, "Specialized image descriptors for signboard photographs classification," in *Databases and Information Systems*, T. Robal, H.-M. Haav, J. Penjam, and R. Matulevičius, Eds. Cham: Springer International Publishing, 2020, pp. 122–129.

[6] A. Samarin and V. Malykh, "Ensemble-based commercial buildings facades photographs classifier," in *Analysis of Images, Social Networks and Texts*, W. M. P. van der Aalst, V. Batagelj, D. I. Ignatov, M. Khachay, O. Koltsova, A. Kutuzov, S. O. Kuznetsov, I. A. Lomazova, N. Loukachevitch, A. Napoli, A. Panchenko, P. M. Pardalos, M. Pelillo, A. V. Savchenko, and E. Tutubalina, Eds. Cham: Springer International Publishing, 2021, pp. 257–265.

[7] V. Malykh and A. Samarin, "Combined advertising sign classifier," in *Analysis of Images, Social Networks and Texts*. Cham: Springer International Publishing, 2019, pp. 179–185.

[8] A. Samarin, A. Savelev, A. Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. A. Motyko, "One-stage classifiers based on u-net and autoencoder with attention for recognition of neoplasms from single-channel monochrome computed tomography images," *Pattern Recognition and Image Analysis*, vol. 33, pp. 132–138, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259310593

[9] M. Ivasic-Kos, M. Pobar, and I. Ipsic, "Automatic movie posters classification into genres," in *ICT Innovations 2014*, A. M. Bogdanova and D. Gjorgjevikj, Eds. Cham: Springer International Publishing, 2015, pp. 319–328.

[10] A. Samarin, A. Savelev, A. Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. Motyko, *Prior Segmentation and Attention Based Approach to Neoplasms Recognition by Single-Channel Monochrome Computer Tomography Snapshots*, 08 2023, pp. 561–570.

[11] A. Samarin, A. Savelev, A.and Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. Motyko, "One-staged attention-based neoplasms recognition method for single-channel monochrome computer tomography snapshots," *Pattern Recognition and Image Analysis*, vol. 32, pp. 645–650, 10 2022.

[12] N. A. Obukhova, A. A. Motyko, U. Kang, S.-J. Bae, and D.-S. Lee, "Automated image analysis in multispectral system for cervical cancer diagnostic," in *2017 20th conference of open innovations association (FRUCT)*. IEEE, 2017, pp. 345–351.

[13] N. Obukhova, A. Motyko, B. Timofeev, and A. Pozdeev, "Method of endoscopic images analysis for automatic bleeding detection and segmentation," in *2019 24th Conference of Open Innovations Association (FRUCT)*. IEEE, 2019, pp. 285–290.

[14] Z. Dai, H. Liu, Q. Le, and M. Tan, "Coatnet: Marrying convolution and attention for all data sizes," *CoRR*, 2021.

[15] A. Radford, J. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," *International Conference on Machine Learning*, 2021.

[16] J. Li, D. Li, C. Xiong, and S. C. H. Hoi, "BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation," *CoRR*, 2022.

[17] J. Li, D. Li, S. Savarese, and S. Hoi, "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," *CoRR*, 2023.

[18] B. Ruan, H.-H. Shuai, and W.-H. Cheng, "Vision transformers: State of the art and research challenges," *CoRR*, 2022.

[19] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 05 2019.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 04 2017.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.

[23] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.

[24] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 248–255.

[25] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *AAAI Conference on Artificial Intelligence*, 02 2016.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[27] M. F. Wahid, T. Ahmed, and M. A. Habib, "Classification of microscopic images of bacteria using deep convolutional neural network," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, 2018, pp. 217–220.

[28] F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 4550–4568, 2018.

[29] L. D. Nguyen, D. Lin, Z. Lin, and J. Cao, "Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–5.

[30] B. Misselwitz, G. Strittmatter, B. Periaswamy, M. Schlumberger, S. Rout, P. Horvath, K. Kozak, and W.-D. Hardt, "Enhanced cell classifier: A multi-class classification tool for microscopy images," *BMC bioinformatics*, vol. 11, p. 30, 01 2010.

[31] R. Liu, W. Dai, T. Wu, M. Wang, S. Wan, and J. Liu, "Aimic: Deep learning for microscopic image classification," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107162, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260722005430

[32] I. G. Shelomentseva, "Classification of light microscopy image using probabilistic bayesian neural network," in *Advances in Neural Computation, Machine Learning, and Cognitive Research VI*, B. Kryzhanovsky, W. Dunin-Barkowski, V. Redko, and Y. Tiumentsev, Eds. Cham: Springer International Publishing, 2023, pp. 265–270.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[34] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[35] M. Yin, Z. Yao, Y. Cao, X. Li, Z. Zhang, S. Lin, and H. Hu, "Disentangled non-local neural networks," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*. Springer, 2020, pp. 191–207.

[36] C. Huang and J. Huang, "A fast hog descriptor using lookup table and integral image," 03 2017.

[37] T. Dittimi and C. Suen, "Modified hog descriptor-based banknote recognition system," *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, 10 2018.

[38] A. Bachchan, A. Gorai, and P. Gupta, "Automatic license plate recognition using local binary pattern and histogram matching," 07 2017, pp. 22–34.

[39] J. Sun, Z. Shisong, and W. Xiaosheng, "Image retrieval based on an improved cs-lbp descriptor," 05 2010, pp. 115 – 117.

[40] O. Tatanov and A. Samarin, "Lfiem: Lightweight filter-based image enhancement model," *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 873–878, 2021.

[41] M. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[42] M. Huang, "Theory and implementation of linear regression," in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2020, pp. 210–217.

[43] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.

[44] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Neural Information Processing Systems*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:3815895

[45] L. F. Scabini and O. M. Bruno, "Structure and performance of fully connected neural networks: Emerging complex network properties," *Physica A: Statistical Mechanics and its Applications*, vol. 615,

p. 128585, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437123001401

[46] "Mmicd dataset." [Online]. Available: https://github.com/itmo-cv-lab/mmicd